1-1-2006

Emotions and Actions Associated with Altruistic Helping and Punishment

Omar Tonsi Eldakar
Binghamton University--SUNY, oeldakar@nova.edu

David Sloan Wilson
Binghamton University--SUNY

Rick O'Gorman
University of Kent at Canterbury - U.K.

Evolutionary Psychology
human-nature.com/ep – 2006. 4: 274-286

Original Article

Emotions and Actions Associated with Altruistic Helping and Punishment

Omar Tonsi Eldakar, Department of Biology, Binghamton University, Binghamton, New York 13902 USA. Email: omar.eldakar@binghamton.edu.

David Sloan Wilson, Department of Biology and Department of Anthropology, Binghamton University, Binghamton, New York 13902 USA.

Rick O'Gorman, Department of Psychology, University of Kent, Canterbury CT2 7NP UK.

Abstract: Evolutionary altruism (defined in terms of fitness effects) exists in the context of punishment in addition to helping. We examine the proximate psychological mechanisms that motivate altruistic helping and punishment, including the effects of genetic relatedness, potential for future interactions, and individual differences in propensity to help and punish. A cheater who is a genetic relative provokes a stronger emotional reaction than a cheater who is a stranger, but the behavioral response is modulated to avoid making the transgression public in the case of cheating relatives. Numerous behavioral differences are not accompanied by emotional differences, suggesting that other psychological mechanisms dictate the specific response to emotion-provoking events. Paradoxically, there is a positive correlation between temptation to cheat and propensity to punish others for cheating, leading to a concept of "selfish punishment" that has been substantiated by a computer simulation model. This study demonstrates that fictional scenarios can provide an important methodological tool for studying the psychological basis of helping and punishment.

Keywords: altruism, punishment, helping, cheating, altruistic punishment, selfish punishment.

Introduction

Altruism and punishment are often contrasted with each other in discussions of human social behavior. An altruist helps others as an end in itself and does not require the threat of punishment. Punishment imposes a cost on non-altruists, making the more selfish alternatives prohibitive.

The recently coined phrase *altruistic punishment* reflects the fact that punishment is often costly for the individual who punishes in addition to the one who is punished (Bowles and Gintis, 2004; Boyd, Gintis, Bowles, and Richerson, 2003). In addition, the benefits of curtailing selfish activities by punishment are often shared by a larger group that includes but is not restricted to the punisher. When these conditions are met, punishment becomes an act of altruism in the evolutionary sense of the word, by increasing the fitness of others at the expense of one's own fitness, or in economic terms by providing a public good at private expense.

To clarify the concept of altruistic punishment, consider a n-person game theory model in which individuals vary for two traits, helping (H) vs. non-helping (NH) and punishing (P) vs.

non-punishing (NP), yielding four strategies: H/P, H/NP, NH/P, and NH/NP. The presence of punishers reduces the fitness of non-helpers and the incidence of non-helping, either by suppressing their behavior or by causing them to leave the group, to the benefit all helpers in the group (not just the punishers). Given these assumptions, it is always the case that H/NP is more fit than H/P within a single group if there is any cost of punishment. Something must be added to the model to make punishment evolutionarily stable, such as between-group selection (Bowles and Gintis, 2004), punishing the non-punishers (which seems to lead to an infinite regress), or conformance cultural transmission (Panchanathan and Boyd, 2004; Fehr, 2004). It is also possible that punishment is evolutionarily unstable in modern social environments, requiring ancestral conditions such as small groups of related individuals that are no longer present (Johnson, Stopka, and Knights, 2003).

These debates about the altruistic nature of punishment are based entirely on fitness effects. Philosophers and psychologists have traditionally defined altruism in terms of motives, or, in evolutionary terms, the proximate psychological mechanisms that motivate behaviors (Sober and Wilson, 1998). From this perspective it is obvious that punishment is motivated by very different psychological mechanisms (e.g., anger and moral outrage) than the helping behaviors typically associated with altruism (e.g., empathy and sympathy).

Recently, O'Gorman, Wilson and Miller (2005) reported an intriguing difference in the psychological response to fictional scenarios that invoke altruistic helping and altruistic punishment. The altruistic helping response was sensitive to genetic relatedness and potential for future interactions, as expected for all forms of altruism (defined in terms of fitness effects) based on kin selection and reciprocity theory. The altruistic punishment response was insensitive to these variables, even though the individuals were clearly motivated to punish in a way that would benefit others at their own expense. O'Gorman et al. measured the psychological response to the fictional scenarios with a relatively small number of questions, such as "How much would you pay to punish the transgressor?" for the punishment scenarios and "How much would you pay to help the person?" in the helping scenarios. In this study, we investigate the psychological mechanisms associated with helping vs. punishment in more detail by having the participants respond to an inventory of words connoting emotion and action. Before proceeding, however, it is important to justify the use of fictional scenarios as an experimental research method.

A comment on methods

Experimental economists have had a large impact on the study of human social behavior, including the concept of altruistic punishment (e.g., Bowles and Gintis, 2004, Fehr 2004). While their interests overlap broadly with those of social and evolutionary psychologists, they tend to adhere to two strong methodological norms; a) never use deception, and b) participants must actually play the games with each other and receive monetary payment. By these standards, experimental economists discount much of the social and evolutionary psychological literature, including the use of fictional scenarios, as methodologically flawed. We think that these norms are unduly restrictive and themselves need to be critiqued. We are not arguing against the study of "real" behavior and monetary payment, of course, but rather for a diversity of research methods. As Robert Putnam (1992, p. 12) put it, "The prudent social scientist, like the wise investor, must rely on diversification to magnify the strengths, and offset the weaknesses, of any single instrument."

The norm against deception is based upon the erosion of trust. Even though a single experiment might gain from the use of deception, it cultivates an attitude of mistrust in the participant population. This is a legitimate problem that needs to be taken seriously by the social scientific community. The norm about "real" behavior and monetary payment is more difficult to justify. The whole thrust of the experimental economics literature is that human behavior cannot be explained by the utility maximizing principle of rational choice theory. People are not driven entirely by monetary concerns but by psychological mechanisms that must be discovered experimentally. Altruistic punishers, for example, are motivated to punish transgressions even at their own monetary expense. If so, then it is ironic to insist that participants will lack motivation unless they receive monetary payment. Participants need to make a good faith effort, which requires motivation, but monetary payment only serves as a means to this end. Monetary payment can even undermine motives to cooperate by transforming a normative situation into a market transaction in the minds of the participants (Mulder, van Dijk, De Cremer, and Wilke, 2006).

As for the necessity of studying "real" behavior, this assumes that mental responses to fictional scenarios and other hypothetical situations are somehow "unreal." Scientists from diverse fields are beginning to recognize the importance of narratives in human psychological and cultural processes. At the neurobiological level, there is evidence that vicarious events are processed by the same circuits as actual events (Bechara, 2002; Berthoz, Armony, Blair, and Dolan, 2002). At the level of individual cognition, narratives are constructed internally to organize experience and rehearse alternative courses of actions (e.g., Bruner, 2002; Cosmides and Tooby 2000; Pennebaker and Seagal 1999). Above the level of the individual, narratives are essential for social transmission and the organization of culture (e.g., Sternberg 1998; Wilson 2002). Even when responses to fictional scenarios do not correspond directly to responses to real-world events, they can reveal psychological mechanisms that motivate real-world behavior in a more complex fashion. The fact that stories engage such interest is a clue to their psychological relevance. It is as if the mind is designed to avidly seek and process information in the form of stories, providing a form of motivation that can equal or exceed monetary payment.

Against this background, it is possible to justify a rigorous scientific methodology based on fictional scenarios, which includes the following elements (see also Wilson and O'Gorman, 2003):

1) Construct a fictional scenario based on the subject of interest (in our case altruistic helping and punishment).

2) Create alternative versions of the scenario that alter key independent variables (in our case, helping vs. punishment, genetic relatives vs. strangers, and potential vs. no potential for future interactions).

3) Measure the response to the independent variables with a number of hypothesis-oriented questions, such as willingness to invest one's own resources to help or punish in our case.

4) In addition to focused questions, we also recommend a more general exploration of emotions and actions elicited by the fictional scenarios (Wilson and O'Gorman, 2003). This can be accomplished by asking one set of participants how they would feel and act, using their open-ended responses to generate a list of emotion and action words, and then measuring the responses of a second set of participants to these words on a numerical scale. This method has the advantage of sampling the emotional and behavioral repertoires of the participant population, including mechanisms and strategies that might not have occurred to the investigators.

5) It is important to include measures of individual differences among the participants. Behavioral responses to experimental economics games are profoundly variable, whether the games are virtual or real. Some individuals cooperate while other cheat; some punish while others refrain. Psychological differences presumably underlie these behavioral differences, which can be investigated with fictional scenarios.

6) Finally, studies involving response to fictional scenarios should be integrated with studies involving response to real-world events, since the relationship between them might well be complex and indirect rather than simple and direct.

To summarize, fictional scenarios easily deserve to be one of the tools in the social scientific toolkit. With these general comments in mind, we can proceed to our particular study.

Methods

The fictional scenario asked the reader to imagine joining an investment club whose members contribute $1000 each to play the stock market. In the punishment version, the investments break even except for a cheater who is discovered to have concealed $200 in profits for himself or herself (the wording was gender neutral). In the helping version, the investments break even and one member is discovered to need help with emergency medical costs. In both versions, genetic relatedness was manipulated by describing the members as cousins vs. strangers. Potential for future interactions was manipulated by describing the cheater or member needing help as still present or having moved to another town (see O'Gorman et al., 2005, for details).

In addition to the basic questions used by O'Gorman et al. (2005), we asked an initial set of volunteer student participants (N=15; 6 females, 9 males) to list words that indicate "how would you feel?" and "how would you act?" in response to the scenarios. Their lists were merged to create the emotion and action words shown in Table 1 for the punishment scenario and Table 2 for the helping scenario. When compiling the word lists, we erred on the side of "splitting" rather than "lumping" since words that seemed synonymous to us might be different in the minds of the participants. These lists can be regarded as the emotional and behavioral repertoire of the participant population, sampled without respect to any particular hypothesis (see Wilson and O'Gorman, 2003, for additional discussion).

The quantitative study was conducted on 330 undergraduate students from Binghamton University's human subject pool (183 females, 126 males, 21 unrecorded) who participated for course research credit as part of a mass testing session. The punishment and helping scenarios and four treatments within each scenario (cousins vs. unrelated and still present vs. moved away) were distributed in random order and each participant responded to a single scenario. After reading the scenario, participants responded to the basic questions and indicated how they would feel and act by responding to the emotion and action words on a scale of 1 (*not at all*) to 9 (*very much*).

Table 1: Initial questions and emotion/action inventory for the punishment scenario.

	Relative	Stranger	Future	No Future	Male	Female	Relative vs. Stranger	Present vs. Absent	Male vs. Female
Basic questions									
How sorry would you feel?	7.23(1.54)	6.77(1.87)	7.08(1.74)	6.93(1.70)	6.84(1.81)	7.16(1.69)			
How likely would you be to help?	7.23(1.58)	6.48(1.98)	7.02(1.70)	6.70(1.94)	6.73(1.90)	6.98(1.83)	**		
What is the most money that you would contribute?	1807(3181)	602(1283)	1286(2449)	1157(2606)	1655(3137)	913(1904)	**		
How angry at those unwilling to help?	5.22(2.44)	5.23(2.18)	5.36(2.17)	5.08(2.45)	4.85(2.33)	5.46(2.32)			
Emotion words									
Apathetic	3.57(2.45)	4.04(2.59)	4.03(2.66)	3.55(2.36)	4.16(2.42)	3.53(2.60)			
Concerned	6.69(2.02)	6.38(2.03)	6.64(2.12)	6.43(1.93)	6.14(2.04)	6.85(2.03)			*
Disdainful	2.89(1.88)	2.93(1.68)	3.05(1.82)	2.77(1.74)	2.93(1.67)	2.78(1.81)			
Empathetic	6.07(2.42)	5.75(2.39)	6.20(2.15)	5.61(2.61)	5.68(2.36)	6.07(2.52)			
Foolish	3.42(2.32)	3.36(1.94)	3.34(2.00)	3.45(2.28)	3.64(2.31)	3.26(2.06)			
Hopeful	6.42(2.34)	6.00(2.35)	6.19(2.59)	6.10(2.31)	6.06(2.48)	6.24(2.52)			
Irresponsible	3.32(2.45)	3.44(2.28)	3.37(2.35)	3.39(2.39)	3.73(2.50)	3.21(2.30)			
Nothing	2.20(1.76)	2.76(2.50)	2.74(2.31)	2.21(1.98)	2.58(1.89)	2.46(2.39)			
Passive	2.65(1.87)	3.45(2.11)	2.86(1.91)	3.23(2.13)	3.24(1.97)	2.93(2.11)	**		
Pity	5.85(2.41)	6.13(2.12)	5.67(2.31)	6.32(2.19)	6.06(2.34)	5.92(2.28)			
Sad	6.28(2.02)	6.12(2.22)	6.08(2.18)	6.33(2.05)	6.31(2.09)	6.14(2.17)			
Scornful	2.61(2.07)	2.50(1.87)	2.52(1.81)	2.59(2.14)	2.72(2.16)	2.43(1.88)			
Sorry	6.55(2.16)	6.30(2.26)	6.18(2.34)	6.70(2.04)	6.52(2.14)	6.37(2.32)			
Sympathetic	6.74(2.25)	6.69(2.10)	6.71(2.25)	6.72(2.11)	6.56(2.20)	6.80(2.21)			
Unconcerned	2.74(2.09)	2.97(2.20)	2.86(1.12)	2.84(2.18)	3.34(2.32)	2.56(2.04)			*
Worried	6.12(2.37)	5.84(2.23)	6.09(2.34)	5.87(2.28)	5.71(2.33)	6.24(2.27)			
Action words									
Act rude	1.70(1.30)	1.75(1.20)	1.67(1.28)	1.77(1.22)	1.91(1.32)	1.60(1.21)			
Act standoffish	2.38(2.11)	3.16(1.97)	2.73(1.88)	2.78(1.86)	2.91(1.87)	2.63(1.85)	**		
Assist in fundraising	6.78(2.38)	5.88(2.41)	6.68(2.18)	6.01(2.38)	6.17(2.42)	6.56(2.17)	*		
By mildly helpful	4.34(1.97)	4.77(2.10)	4.45(2.17)	4.65(2.35)	4.96(2.30)	4.18(2.16)			*
Express sympathy	7.37(2.06)	7.10(2.02)	7.23(1.93)	7.25(2.08)	7.00(2.20)	7.40(1.86)			
Find out expenses	6.92(1.51)	5.90(2.47)	6.75(2.18)	6.11(2.42)	6.58(2.31)	6.39(2.33)	**		
Help if very serious	8.07(2.19)	7.36(1.92)	7.71(1.61)	7.74(1.89)	7.71(1.87)	7.76(1.70)	*		
Lend money	6.52(2.01)	5.62(2.30)	6.21(2.07)	5.74(2.44)	5.88(2.26)	6.03(2.30)			
Not bring it up	3.07(2.69)	3.76(2.09)	3.25(1.92)	3.55(2.22)	3.66(2.21)	3.20(2.01)	*		
See if they ask for help	4.90(2.69)	5.50(2.08)	4.93(2.39)	5.45(2.46)	5.03(2.25)	5.16(2.58)			
Try to help	7.32(1.75)	6.44(2.30)	7.06(2.00)	6.73(2.14)	6.88(2.19)	6.97(2.01)	**		

Note: Rows indicate the initial survey questions and emotion and action words. Columns indicate the mean value (with standard deviation in parentheses) of dollar amounts (for rows 3-4) or agreement on a 9 point scale (for all other rows) when the cheater was a relative vs. a stranger, presence vs. absence of potential for future interactions, and the response of male and female participants. The final three columns indicate the main effects of an ANOVA (* = <.05, ** = <.01, *** = <.001, **** = <.0001). Full statistical results are available upon request.

Table 2: Initial questions and emotion/action inventory for helping scenario.

	Relative	Stranger	Future	No Future	Male	Female	Relative vs. Stranger	Present vs. Absent	Male vs. Female
Basic questions									
How sorry would you feel?	7.23(1.54)	6.77(1.87)	7.08(1.74)	6.93(1.70)	6.84(1.81)	7.16(1.69)			
How likely would you be to help?	7.23(1.58)	6.48(1.98)	7.02(1.70)	6.70(1.94)	6.73(1.90)	6.98(1.83)	**		
What is the most money that you would contribute?	1807(3181)	602(1283)	1286(2449)	1157(2606)	1655(3137)	913(1904)	**		
How angry at those unwilling to help?	5.22(2.44)	5.23(2.18)	5.36(2.17)	5.08(2.45)	4.85(2.33)	5.46(2.32)			
Emotion words									
Apathetic	3.57(2.45)	4.04(2.59)	4.03(2.66)	3.55(2.36)	4.16(2.42)	3.53(2.60)			
Concerned	6.69(2.02)	6.38(2.03)	6.64(2.12)	6.43(1.93)	6.14(2.04)	6.85(2.03)			*
Disdainful	2.89(1.88)	2.93(1.68)	3.05(1.82)	2.77(1.74)	2.93(1.67)	2.78(1.81)			
Empathetic	6.07(2.42)	5.75(2.39)	6.20(2.15)	5.61(2.61)	5.68(2.36)	6.07(2.52)			
Foolish	3.42(2.32)	3.36(1.94)	3.34(2.00)	3.45(2.28)	3.64(2.31)	3.26(2.06)			
Hopeful	6.42(2.34)	6.00(2.35)	6.19(2.59)	6.10(2.31)	6.06(2.48)	6.24(2.52)			
Irresponsible	3.32(2.45)	3.44(2.28)	3.37(2.35)	3.39(2.39)	3.73(2.50)	3.21(2.30)			
Nothing	2.20(1.76)	2.76(2.50)	2.74(2.31)	2.21(1.98)	2.58(1.89)	2.46(2.39)			
Passive	2.65(1.87)	3.45(2.11)	2.86(1.91)	3.23(2.13)	3.24(1.97)	2.93(2.11)	**		
Pity	5.85(2.41)	6.13(2.12)	5.67(2.31)	6.32(2.19)	6.06(2.34)	5.92(2.28)			
Sad	6.28(2.02)	6.12(2.22)	6.08(2.18)	6.33(2.05)	6.31(2.09)	6.14(2.17)			
Scornful	2.61(2.07)	2.50(1.87)	2.52(1.81)	2.59(2.14)	2.72(2.16)	2.43(1.88)			
Sorry	6.55(2.16)	6.30(2.26)	6.18(2.34)	6.70(2.04)	6.52(2.14)	6.37(2.32)			
Sympathetic	6.74(2.25)	6.69(2.10)	6.71(2.25)	6.72(2.11)	6.56(2.20)	6.80(2.21)			
Unconcerned	2.74(2.09)	2.97(2.20)	2.86(1.12)	2.84(2.18)	3.34(2.32)	2.56(2.04)			*
Worried	6.12(2.37)	5.84(2.23)	6.09(2.34)	5.87(2.28)	5.71(2.33)	6.24(2.27)			
Action words									
Act rude	1.70(1.30)	1.75(1.20)	1.67(1.28)	1.77(1.22)	1.91(1.32)	1.60(1.21)			
Act standoffish	2.38(2.11)	3.16(1.97)	2.73(1.88)	2.78(1.86)	2.91(1.87)	2.63(1.85)	**		
Assist in fundraising	6.78(2.38)	5.88(2.41)	6.68(2.18)	6.01(2.38)	6.17(2.42)	6.56(2.17)	*		
By mildly helpful	4.34(1.97)	4.77(2.10)	4.45(2.17)	4.65(2.35)	4.96(2.30)	4.18(2.16)			*
Express sympathy	7.37(2.06)	7.10(2.02)	7.23(1.93)	7.25(2.08)	7.00(2.20)	7.40(1.86)			
Find out expenses	6.92(1.51)	5.90(2.47)	6.75(2.18)	6.11(2.42)	6.58(2.31)	6.39(2.33)	**		
Help if very serious	8.07(2.19)	7.36(1.92)	7.71(1.61)	7.74(1.89)	7.71(1.87)	7.76(1.70)	*		
Lend money	6.52(2.01)	5.62(2.30)	6.21(2.07)	5.74(2.44)	5.88(2.26)	6.03(2.30)			
Not bring it up	3.07(2.69)	3.76(2.09)	3.25(1.92)	3.55(2.22)	3.66(2.21)	3.20(2.01)	*		
See if they ask for help	4.90(2.69)	5.50(2.08)	4.93(2.39)	5.45(2.46)	5.03(2.25)	5.16(2.58)			
Try to help	7.32(1.75)	6.44(2.30)	7.06(2.00)	6.73(2.14)	6.88(2.19)	6.97(2.01)	**		

Note: Rows indicate the initial survey questions and emotion and action words. Columns indicate the mean value (with standard deviation in parentheses) of dollar amounts (for row 3) and agreement on a 9 point scale (for all other rows) when the person needing help was a relative vs. a stranger, presence vs. absence of potential for future interactions, and the response of male and female participants. The final three columns indicate the main effects of an ANOVA (* = <.05, ** = <.01, *** = <.001, **** = <.0001). Full statistical results are available upon request.

Results

For the questions that they share in common, the results of this study are largely consistent with those of O'Gorman et al. (2005), as shown for the punishment scenario in Figure 1. It is important to show the full distribution of responses because they are highly non-normal and informative in their own right. The most frequent response to question 1 ("how angry would you feel toward this person?") was the maximum response of 9 (Figure 1A). The desire to punish was also strong, peaking at a value of 7 (Figure 1B). Most participants thought that the cheater should pay back what was stolen (the modal value is $200) but some indicated much more (Figure 1C). Despite the anger and desire to punish reported in questions 1 and 2, the amount that the subjects were willing to pay to punish was highly variable (Figure 1D), including 54% who were unwilling to pay anything and a few individuals who indicated willingness to pay exorbitant amounts. Anger toward non-punishers was more muted than toward the cheater but still reached high values for some individuals (Figure 1E). Finally, most participants felt that they would not be tempted to cheat in this situation but a sizeable proportion was tempted to varying degrees (Figure 1F).

These results indicate that the fictional scenarios were successful at engaging a strong psychological response from the participants. The surprising result of O'Gorman et al. (2005) was that *the response was not influenced by genetic relatedness or potential for future interaction*, at least in terms of the basic questions listed in rows 1-6 of Table 1. This result must be revised for genetic relatedness, although not for future interactions, on the basis of the present study. Not only was there a minor effect of genetic relatedness for the basic question "How angry would you feel?", but numerous effects were revealed by the inventory of emotion and action words. With respect to emotions, in addition to feeling more angry, participants also felt more betrayed, disappointed, disgusted, hurt, and shocked at cheating by a cousin than by a stranger. With respect to action, participants indicated a greater likelihood of confronting and seeking an apology from a relative than a stranger, they but were more likely to warn others of a stranger than a relative. With respect to sex differences, men indicated a greater willingness to demand money, ostracize, physically hurt, threaten, and yell at the cheater than women.

The distribution of responses to the first four questions of the helping scenario are shown in Figure 2, indicating a strong empathetic response (2A) and desire to help (2B), a highly skewed distribution in the amount willing to pay (2C), and the full spectrum of anger toward others not willing to help (2D). The first four lines of Table 2 show that genetic relatedness strongly influences desire to help and amount of money contributed (similar to the previous study) but no effect of potential for future interactions (unlike the previous study). The inventory of emotion and action words enable the psychological response to helping to be examined in more detail. Surprisingly, the effect of genetic relatedness did *not* take place at the level of emotional response (with the exception of the word "passive") but primarily at the level of behavioral response. Participants reported a greater willingness to help cousins than strangers, but not because they had a greater emotional response to their dilemma. With respect to sex differences females exhibited greater concern and willingness to help than males.

Figure 1. Histograms indicating dollar amounts (1C and 1D) and agreement on a nine-point scale (all other graphs) to the basic questions of the punishment scenario.

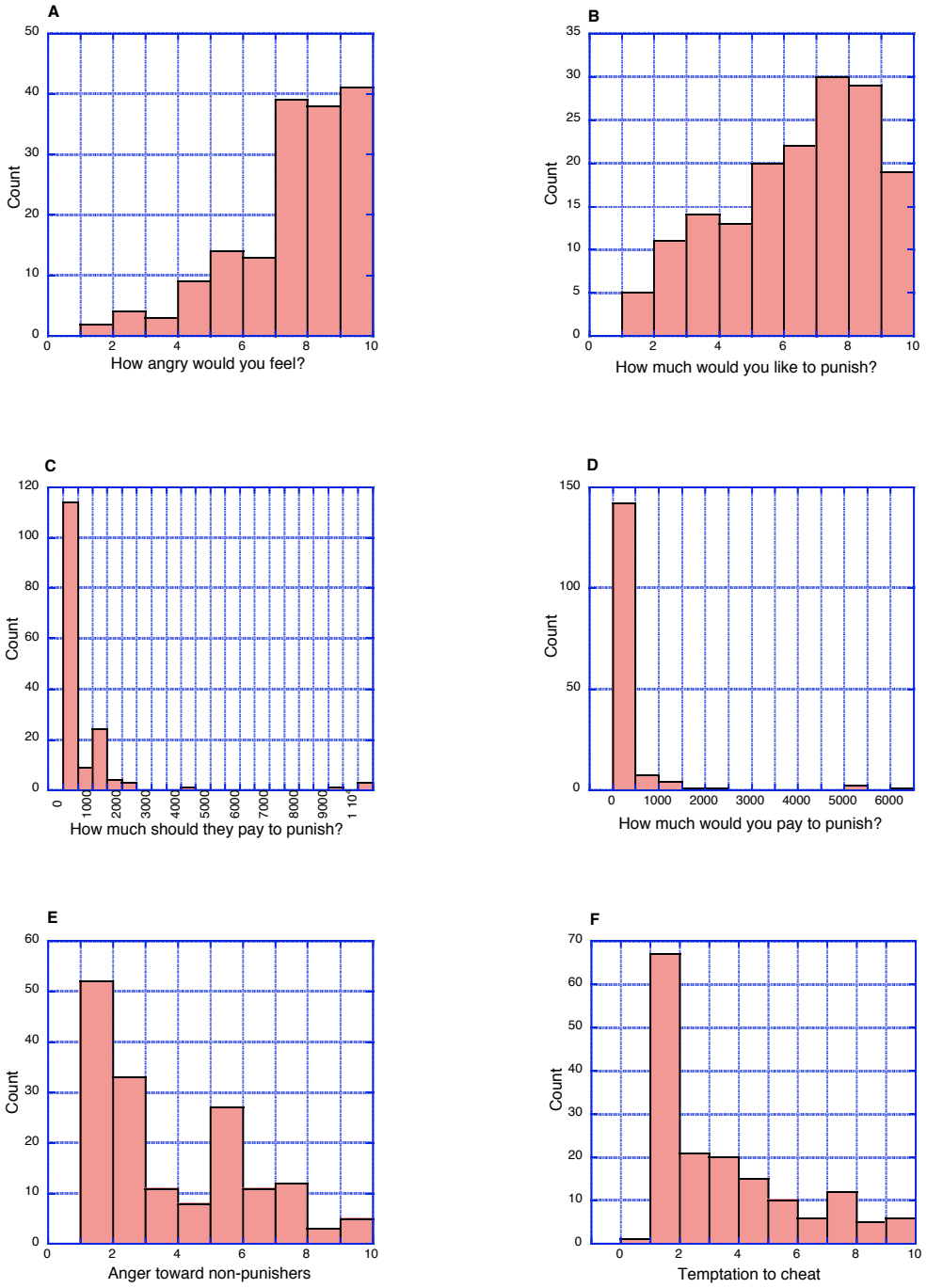
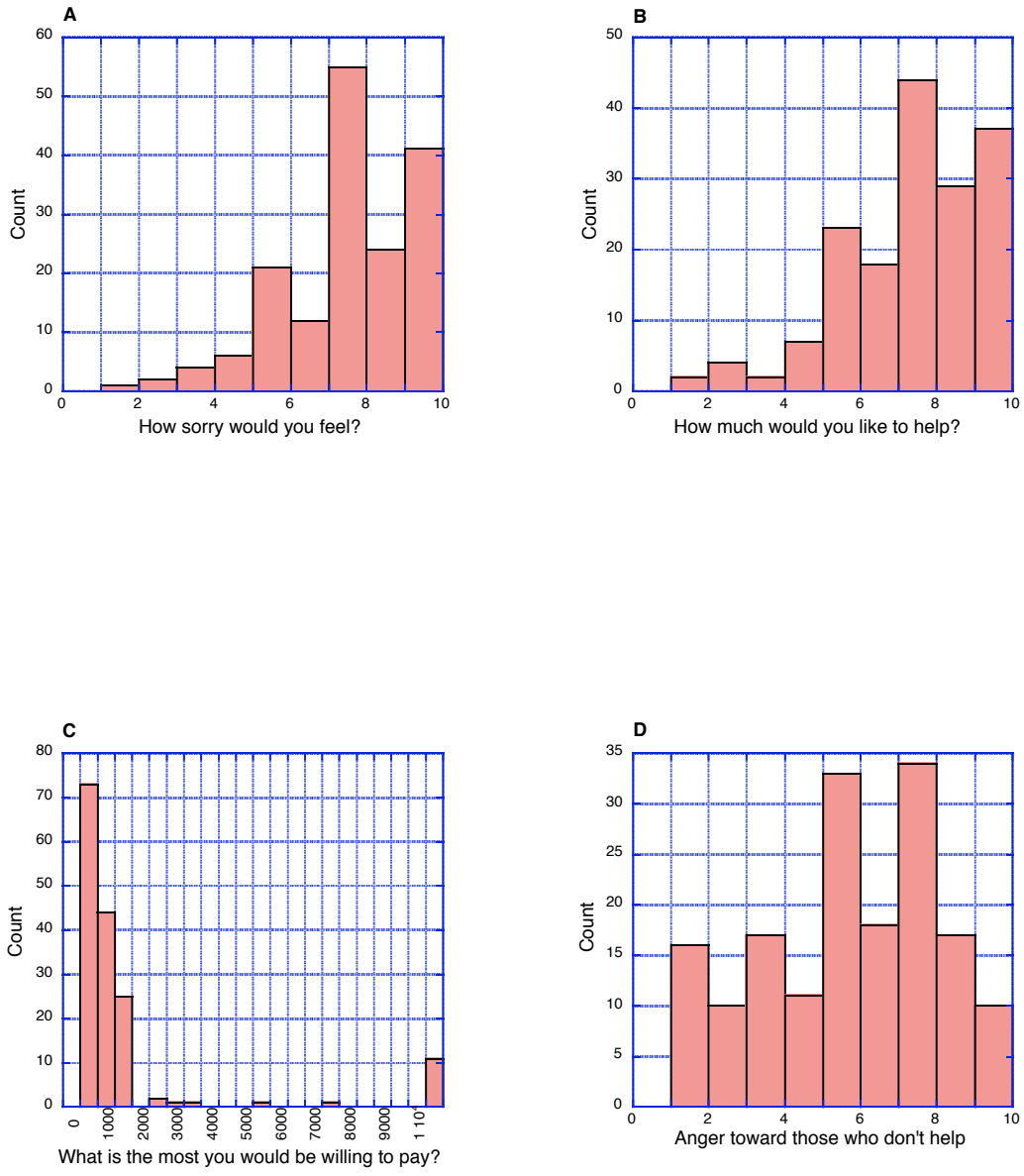


Figure 2. Histograms indicating dollar amounts (2C) and agreement on a nine-point scale (all other graphs) to the basic questions of the helping scenario.



Individual differences in altruistic punishment and helping

To examine individual differences in willingness to punish at one's own expense, we used the item "What is the most you would be willing to pay to punish this person?" as the dependent variable in three stepwise multiple regression analyses. In the first, the other five basic questions were used as the independent variables. The total R^2 was .26 (df=153, F=14.7, p<.0001) with four variables retained in the following order (parentheses report the t-ratio, significance value, and incremental effect on the total R^2 value): How much the cheater should pay (t=5.10, p<.0001, DR^2=.16), the subject's temptation to cheat (t=2.29, p=.0007, DR^2=.06), anger toward non-punishers (t=2.82, p=.0374, DR^2=.02), and anger toward the cheater (t=-2.58, p=.011, DR^2=.03). The second and fourth variables are counterintuitive; evidently, willingness to pay to punish the cheater is *positively* related to one's own temptation to cheat and *negatively* related to anger toward the cheater.

The second analysis used the emotion words as the independent variables. Surprisingly, none of them were significantly related to the dependent variable. The third analysis used the action words as the dependent variables. The only significant variable was "physically hurt them" (R^2=.09, df=149, F=15.8, p<.0001).

Sex differences in altruistic punishment were examined by dividing the responses into three categories: willing to pay nothing (n=80), willing to pay between 1-99 dollars (n=30), and willing to pay greater than 100 dollars (n=37). Males are more frequent in the third category but the difference is not statistically significant (X^2=4.75, df=2, p=.09).

A similar set of analyses was conducted for altruistic helping using the question "What is the most you would be willing to give?" as the dependent variable. When the other basic questions were used as the independent variables, the total R^2 was *.13* (df=156, F=12.2, p<.0001) with two variables retained in the following order: How much they would like to help the person (t=4.93, p<.0001, DR^2=.10), and a negative correlation with anger toward non-helpers (t=-2.24, p=.0265, DR^2=.02). In the second analysis, none of the emotion words were related. In the third analysis, the only significant action variable was "help if very serious" (R^2=.04, df=151, F=6.94, p<.01). There was no sex difference in the amount that participants were willing to give (X^2=1.43, df=2, p=.49).

Discussion

The most notable result of the O'Gorman et al. study (2005) was that altruistic punishment is insensitive to genetic relatedness and potential for future interactions, in contrast to altruistic helping. The current study used an inventory of emotion and action words to provide a better indicator of psychological mechanisms. Our results show that participants are sensitive to genetic relatedness in their response to a social transgression, in contrast to the earlier study. A cheating cousin provokes a greater negative emotional reaction than a cheating stranger. However, the more negative emotional response does not lead directly to a more negative behavioral response. Instead, the behavioral response is modulated in a way that is more negative toward relatives in some respects (confronting and seeking an apology) and more negative toward strangers in other respects (warning others about the transgression).

Another complex relationship between emotions and actions exist for altruistic helping. At the emotional level, participants responded equally to the misfortune of relative and a stranger, but the undifferentiated emotional response led to very different actions; active helping

in the case of relatives and avoidance in the case of strangers. Having the participants respond to an inventory of emotion and action words is a simple and effective technique for identifying these complex relationships.

No consistent pattern emerges for the potential for future interactions. O'Gorman et al. (2005) showed no effect for punishment in two replicated experiments and a strong effect for helping in a single experiment. However, we show no effect for any of our treatments in both helping and punishing scenarios, a discrepancy that will require future studies to resolve.

Both studies reveal a response to cheating that at first seems paradoxical: Most participants are very angry and think that the cheater should be punished, but over half are unwilling to pay even a penny to accomplish the job. These individuals are not necessarily hypocritical, however, because human life affords many ways to punish cheaters at little or no personal expense, such as gossiping to undermine their reputation (Wilson, Wilczynski, Wells, and Weiser, 2000) or avoiding future interactions (Wilson and Dugatkin, 1997). These forms of punishment can provide a potent outlet for individuals who are very angry but unwilling to invest their own resources.

The remaining individuals indicate a stronger commitment to "put their money where their mouth is" by punishing cheaters at their own expense, in amounts ranging from modest to far greater than what they ($20) or anyone else ($200) lost, as if punishing transgressions is a matter of principle (Frank, 1988; Hirshleifer, 1987). The commitment to punish at one's own expense is not a simple function of emotional intensity. In fact, there is a weak *inverse* correlation between the degree of anger and the amount that participants were willing to pay in the first multiple regression analysis. The strongest correlations were with how much the cheater should pay, one's own temptation to cheat, and anger toward others who don't punish.

A positive correlation between temptation to cheat and willingness to punish other cheaters is one of the most interesting results of this study. It raises the possibility that "altruistic punishment" is in fact selfish, a strategy employed by cheaters to eliminate competition from other cheaters. Eldakar, Farrell, and Wilson (2006) have investigated this possibility in a computer simulation model that emulates a public goods situation similar to our fictional scenario and experimental economics games. Even when helping and punishment are initially uncorrelated traits, a positive correlation between cheating and propensity to punish other cheaters robustly emerges from the model, suggesting that our empirical result might prove to be quite general.

The psychological mechanisms the lead to helping and punishment are complex, including emotional responses that are modulated by other mechanisms to produce behaviors appropriate to given situations and people. Anger can lead to avoidance, aggression, diplomacy, or inaction if the costs are too great. Sympathy can lead to direct helping (such as stopping to aid an accident victim), indirect helping (such as calling 911), or inaction (such as emotionally responding to a charitable appeal but failing to give money). A combination of methods will be required to fully understand these mechanisms. This study shows that fictional scenarios provide a useful tool that can be used in combination with other tools, such as computer simulation models and studies of actual behavior.

Acknowledgements: We thank Ralph Miller, Anne B. Clark, N. Prince and other members of EvoS, Binghamton University's evolutionary studies program, for advice and discussion on this project. Supported by a grant to DSW from the Institute on Research on Unlimited Love.

Received 1 July 2006; Revision received 19 July 2006; Accepted 25 July 2006

References

Bechara, A. (2002). The neurology of social cognition. *Brain, 125,* 1673-1675.

Berthoz, S., J. L. Armony, R. J. R. Blair, and R. J. Dolan. (2002). An fMRI study of intentional and unintentional (embarrassing) violations of social norms. *Brain, 125,* 1696-1708.

Bowles, S., and Gintis, H. (2004). The evolution of strong reciprocity: cooperation in heterogeneous populations. *Theoretical Population Biology, 65,* 17-28.

Boyd, R., Gintis, H., Bowles, S., and Richerson, P. J. (2003). The evolution of altruistic punishment. *Proceedings of the National Academy of Sciences USA, 100,* 3531–3535.

Bruner, J. (2002). *Making stories: Law, literature, life.* New York: Farrar, Straus, and Giroux.

Cosmides, L., and Tooby, J. (2000). Evolutionary psychology and the emotions, In M. Lewis, and J. M. Haviland-Jones, (Eds.), *Handbook of emotions* (pp. 91-115). New York: Guilford press.

Eldakar, O.T., D.L. Farrell, and D.S. Wilson (2006). *Selfish punishment: altruism can be maintained by competition among cheaters.* Manuscript under editorial review.

Fehr, E. (2004). Don't lose your reputation. *Nature, 432,* 449-450.

Frank, R. H. (1988). *Passions within reason.* New York: W.W. Norton.

Hirshleifer, J. and J. C. M. Col (1988). What strategies can support the evolutionary emergence of cooperation? *Journal of Conflict Resolution, 332,* 367-398.

Johnson, D.D.P., Stopka, P., and Knights, S. (2003). The puzzle of human cooperation. *Nature, 421,* 911-912.

Mulder, L. B., E. van Dijk, D. De Cremer, and H. A. M. Wilke, (2006). Undermining trust and cooperation: The paradox of sanctioning systems in social dilemmas. *Journal of Experimental Social Psychology, 42,* 147-162.

O'Gorman, R., Wilson, D. S., and Miller, R. (2005). Altruistic punishment and helping differ in sensitivity to relatedness, friendship, and future interactions. *Evolution and Human Behavior, 26,* 375-387

Panchanathan, K., and Boyd, R. (2004). Indirect reciprocity can stabilize cooperation without the second-order free rider problem. *Nature, 432,* 499-502.

Pennebaker, J. W., and Seagal, J. D. (1999). Forming a story: the health benefits of narrative. *Journal of Clinical Psychology, 55,* 1243-1254.

Putnam, R. D. (1992). *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton, NJ: Princeton University Press.

Sober, E., and Wilson, D. S. (1998). *Unto others: The evolution and psychology of unselfish behavior*. Cambridge, MA: Harvard University Press.

Sternberg, R. J. (1998). *Love is a story: A new theory of relationships.* New York: Oxford University Press.

Wilson, D. S. (2002). *Darwin's Cathedral: Evolution, religion, and the nature of society*. Chicago: University of Chicago Press.

Wilson, D. S., and Dugatkin, L. A. (1997). Group selection and assortative interactions. *American Naturalist, 149,* 336-351.

Wilson, D. S. and O'Gorman Rick (2003). Emotions and actions associated with norm-breaking Events. *Human Nature, 14,* 277-304.

Wilson, D. S., Wilczynski, C., Wells, A., and Weiser, L. (2000). Gossip and other aspects of language as group-level adaptations. In C. Heyes and L. Huber. (Eds.) *Cognition and Evolution* (pp. 347-365). Cambridge, MA: MIT Press.